# Module 11 - Probability and Statistics

ME3023 - Measurements in Mechanical Systems

Mechanical Engineering

Tennessee Technological University

## Topic 4 - Statistics of Finite-Sized Data Sets

## Topic 4 - Statistics of Finite-Sized Data Sets

- our example continues...
- the entire population
- a finite-sized data set
- Student's t Distribution

# our example continues...

as we have discussed, we cannot measure all of the ball bearings, we have a fixed-size data set

# the entire population

The true variance is:

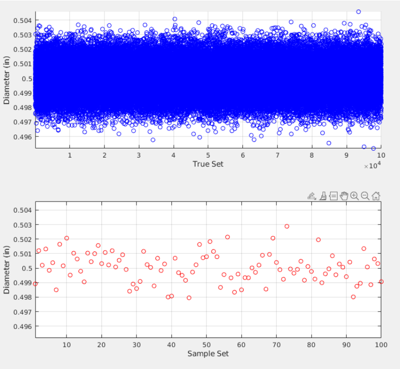
$$\sigma^2 = \int_{-\infty}^{\infty} (x - x')^2 p(x) dx$$

For discrete data this becomes:

$$\sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - x')^2$$

The square root of the variance is the standard deviation.

$$\sigma = \sqrt{\sigma^2} = sqrt \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - x')^2$$



# a finite-sized data set

We use a similar set of statistical parameters to describe the population after the true population has been sampled.

The sample mean is:

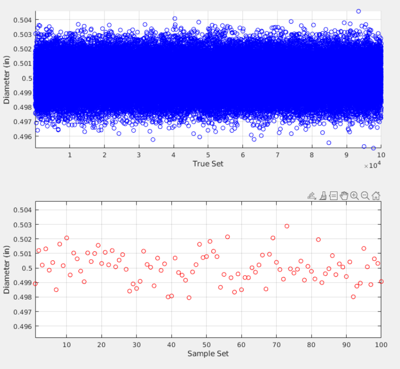
$$\bar{x} = \sum_{i=1}^{N} x_i$$

The sample variance and sample standard deviation:

$$s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

The square root of the sample variance is the sample standard deviation.

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$



# a finite-sized data set

The sample mean value provides a most probable estimate of the true mean value, $x'$. The sample variance represents a probable measure of the variation found in a data set. The degrees of freedom, $n$, in a statistical estimate equate to the number of data points minus the number of previously determined statistical parameters used in estimating that value. For example, the degrees of freedom in the sample variance is $\nu = N - 1$, as seen in denominator of Equations 4.14b and c. These equations are robust and are used regardless of the actual probability density function of the measurand.

# a finite-sized data set

The relation between probability and infinite statistics can be extended to data sets of finite sample size with only some modification. When data sets are finite or smaller than the population, the $z$ variable does not provide a reliable weight estimate of the true probability. However, the sample variance can be weighted in a similar manner so as to compensate for the difference between the finite statistical estimates and the statistics based on an assumed $p(x)$. For a normal distribution of $x$ about some sample mean value, $x$, we can state that statistically

$$x_i = \bar{x} \pm t_{\nu,P} s_x \quad P(\%)$$

where the variable t n,P provides a coverage factor used for finite data sets and which replaces the z variable. This new variable is referred to as the StudentâĂŹs t variable,

$$t = \frac{\bar{x} - x'}{s_x/\sqrt{N}}$$

The interval $\pm t_{\nu,P} s_x$ represents a precision interval, given at probability $P\%$, within which one should expect any measured value to fall.

# Student's t Distribution

**Table 4.4** Student's $t$ Distribution

| $\nu$ | $t_{50}$ | $t_{90}$ | $t_{95}$ | $t_{99}$ |
|---|---|---|---|---|
| 1 | 1.000 | 6.314 | 12.706 | 63.657 |
| 2 | 0.816 | 2.920 | 4.303 | 9.925 |
| 3 | 0.765 | 2.353 | 3.182 | 5.841 |
| 4 | 0.741 | 2.132 | 2.770 | 4.604 |
| 5 | 0.727 | 2.015 | 2.571 | 4.032 |
| 6 | 0.718 | 1.943 | 2.447 | 3.707 |
| 7 | 0.711 | 1.895 | 2.365 | 3.499 |
| 8 | 0.706 | 1.860 | 2.306 | 3.355 |
| 9 | 0.703 | 1.833 | 2.262 | 3.250 |
| 10 | 0.700 | 1.812 | 2.228 | 3.169 |
| 11 | 0.697 | 1.796 | 2.201 | 3.106 |
| 12 | 0.695 | 1.782 | 2.179 | 3.055 |
| 13 | 0.694 | 1.771 | 2.160 | 3.012 |
| 14 | 0.692 | 1.761 | 2.145 | 2.977 |
| 15 | 0.691 | 1.753 | 2.131 | 2.947 |
| 16 | 0.690 | 1.746 | 2.120 | 2.921 |
| 17 | 0.689 | 1.740 | 2.110 | 2.898 |
| 18 | 0.688 | 1.734 | 2.101 | 2.878 |
| 19 | 0.688 | 1.729 | 2.093 | 2.861 |
| 20 | 0.687 | 1.725 | 2.086 | 2.845 |
| 21 | 0.686 | 1.721 | 2.080 | 2.831 |
| 30 | 0.683 | 1.697 | 2.042 | 2.750 |
| 40 | 0.681 | 1.684 | 2.021 | 2.704 |
| 50 | 0.680 | 1.679 | 2.010 | 2.679 |
| 60 | 0.679 | 1.671 | 2.000 | 2.660 |
| $\infty$ | 0.674 | 1.645 | 1.960 | 2.576 |